

University of Groningen

Are musicians at an advantage when processing speech on speech?

Kaplan, Elif Canseza; Wagner, Anita; Başkent, Deniz

Published in:
Proceedings of ICMPC15/ESCOM10

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kaplan, E. C., Wagner, A., & Başkent, D. (2018). Are musicians at an advantage when processing speech on speech? In *Proceedings of ICMPC15/ESCOM10* (pp. 233-236). Department of Musicology, University of Graz.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Are musicians at an advantage when processing speech on speech?

Elif Canseza Kaplan,^{*#1} Anita Wagner,^{*#2} and Deniz Başkent^{*#3}

^{*} *Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Netherlands*

[#] *Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Netherlands*

¹*e.c.kaplan@rug.nl*, ²*a.wagner@umcg.nl*, ³*d.baskent@umcg.nl*

Abstract

Several studies have shown that musicians may have an advantage in a variety of auditory tasks, including speech in noise perception. The current study explores whether musical training enhances understanding two-talker masked speech. By combining an off-line and an on-line measure of speech perception, we investigated how automatic processes can contribute to the potential perceptual advantage of musicians. Understanding the underlying mechanisms for how musical training may lead to a benefit in speech-in-noise perception could help clinicians in developing ways to use music as a means to improve speech processing of hearing impaired individuals.

Introduction

Earlier studies have shown that musical training can grant normal-hearing listeners an advantage on auditory tasks, in particular for speech comprehension in noise or in the presence of background talkers (Başkent & Gaudrain, 2016; Kraus, Strait, & Parbery-Clark, 2012; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Swaminathan et al., 2015). However, not all studies have found a difference between musicians' and non-musicians' speech in noise performance (Boebinger et al., 2015; Ruggles, Freyman, & Oxenham, 2014), or they have found an advantage in some auditory tasks and not in others (Fuller, Galvin, Maat, Free, & Başkent, 2014). Taken together, various studies addressing a musician advantage provided mixed results (for a review see: Coffey, Mogilever, & Zatorre, 2017), which can be partly explained through the use of different measures (e.g. behavioral versus physiological, online versus offline), the variations in target/noise signal properties, and the linguistic complexity of the target and noise stimuli used in the tasks.

In the current study, we aimed to investigate how perception of speech in two-talker masker noise differs between musicians and non-musicians. We first used a sentence-recall task, in which participants recall and repeat target Dutch sentences presented with two-talker Dutch sentence maskers. This task provides an estimate of sensory sensitivity in different Target-Masker-Ratios (TMR) by estimating the percentage of correctly recalled words. This is an offline-measure of speech processing, similar to the single-talker masker study reported by Başkent & Gaudrain (2016). Following, we used a visual-world paradigm, using eye tracking (Cooper, 1974; for a review see: Salverda & Tanenhaus, 2017). Here, while listening to sentences, participants visually search for the image of the target word mentioned by the target speaker. The target image is displayed among a competitor (phonologically similar to the target, i.e. ham-hamster) and two distractor images. During the task, gaze fixations and pupillary responses are measured. This task

gives an online measure of speech processing by comparing the time course of gaze fixations to the target and/or the competitor word, as well as the changes in event related pupil dilations. The online measure indicates how quickly participants integrate the acoustic information in the signal when they are accessing the mental lexicon and the extent of the mental effort involved in processing of linguistic information. We used the utterances of the same target speaker and the masker speaker in both tasks in order to be able to capture how the automatic processes may contribute to speech on speech perception in musicians.

Our main research questions are:

1. Would the difference observed in an offline task, manifest itself in an online task?
 - Musicians have been previously shown to perform better than non-musicians in single talker masker (Başkent & Gaudrain, 2016) or four-talker babble (Slater & Kraus, 2016). Based on literature, we expect musicians to be less affected by masked speech than non-musicians on the offline task.
2. Would the results obtained by an online task be clearer in reflecting the difference in the usage of cognitive resources in musicians and non-musicians?
 - Speech masker should have a smaller effect in terms of timing of lexical mapping and pupil dilations for musicians than non-musicians

Method

Participants. Ten musicians (years of musical training = 10.6, age = 23.8) and twenty non-musicians (years of musical training = 1.63, age = 25.1) from Groningen, the Netherlands were recruited for participating in the study. The musicianship criteria were taken from the literature (Fuller et al., 2014) and included the following: having started music before the age of 7, having at least 10 years of musical training, actively practicing music at least for 3 years prior to the study. The non-musician criteria were not meeting the musician criteria and not having more than 3 years of music training.

Materials. The same target and masker speakers' sentences were utilized in both tasks.

In the sentence recall task, 28 semantically neutral Dutch target sentences (Wagner, Toffanin, & Baskent, 2016) uttered by a female speaker were embedded in two-talker maskers. We chose to use a two-talker masker based on previous literature. According to Rosen, Souza, Ekelund, & Majeed (2013), greatest changes in masking occur when the number of talkers change from 1 to 2 or 4 background talkers, and Calandruccio, Buss, & Bowdrie (2017) showed that a two-

talker masker was the most effective masker. The masker sentences consisted of simple, meaningful Dutch sentences (Versfeld, Daalder, Festen, & Houtgast, 2000) uttered by a different female speaker than the target speaker.

The onset of the target sentences was delayed by 500 ms from the maskers' onset, and the participants were instructed to repeat the sentence that started later. The sentences were presented at four possible TMRs in a random order; -3, -5, -7 or -9 dB SPL (Calandruccio et al., 2017), where the presentation level of the masker sentence combination was fixed at 75 dB SPL and the presentation level of the target sentence was adjusted depending on the TMR in each condition.

Based on pilot results of the sentence recall task, the eye-tracking experiment's TMR levels were set to 0 and -5 dB SPL. The stimuli were presented either in a two-talker masker with 0 or -5 TMR and, as a baseline in quiet (without any masking).

Procedure. Participants were first tested for normal hearing, where the hearing levels were < 20 dB HL for pure tone thresholds from 250 to 4000 Hz bilaterally, then first completed the visual world paradigm, and following the sentence recall task.

In the visual world paradigm, the experiment started with a practice phase where participants completed 8 trials of the quiet and masked conditions. They were instructed to pay attention to the voice they heard in the quiet condition throughout the experiment and choose the image of the target word from four images displayed on the screen. The quiet condition was presented first at all times. Upon completing the quiet condition, participants could take a break and then complete the masked condition. The two TMR levels were presented in a random order within one block. Each condition included 12 trials. The order of presentation of the sentences was randomized for each participant.

The sentence recall task began with a practice phase of 4 trials, where participants heard an example of each TMR level (-3, -5, -7, -9 dB). They were instructed to repeat the sentence uttered by the same target speaker as in the first experiment. After the practice, participants completed 28 trials, with TMR conditions presented randomly in order to eliminate learning effects.

Results

The preliminary results will be reported; however, none are conclusive since the musician data is based on 10 participants.

In the sentence recall task, musicians perform slightly better as the level of background masker speech increases (Figure 1).

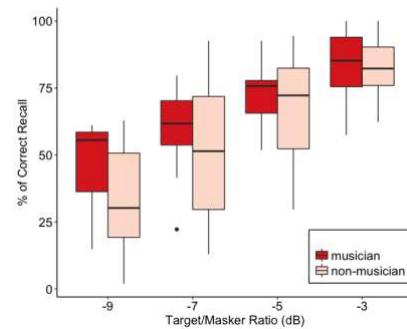


Figure 1. Percentage of correctly recalled keywords as a function of target masker ratios (-9, -7, -5, -3 dB) of the sentence recall task.

The gaze fixations in the quiet condition (Figure 2) indicate that lexical competition occurs in a similar fashion both for musicians and non-musicians. The point of disambiguation is happening earlier in quiet than in masked conditions for both groups. The proportion of fixation to the target is also higher for both groups in quiet versus in masked speech.

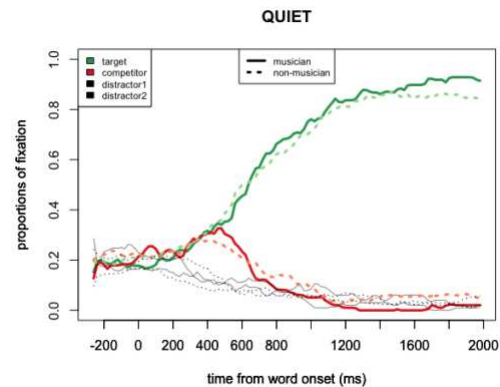


Figure 2. Gaze fixations to the target, competitor and distractor images as a function of time from word onset of the target word in the quiet condition.

At TMR=0 dB, where the intensity of the masker and the target are equal, the point of disambiguation seems to be slightly delayed for non-musicians. The lexical competition is still present for both groups, while the certainty is lowered for both groups as shown by the proportion of fixation to the target (Figure 3).

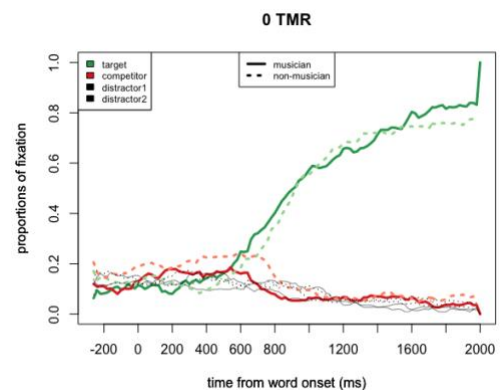


Figure 3. Gaze fixations to the target, competitor and distractor images as a function of time from word onset of the target word in TMR=0 dB condition.

In the most difficult TMR, where the target is presented at 5 dB lower level than the masker the lexical competition is less present for musicians, but not for non-musicians. Overall fixations to the target are reduced compared to the other conditions for both groups, indicating increased uncertainty (Figure 4).

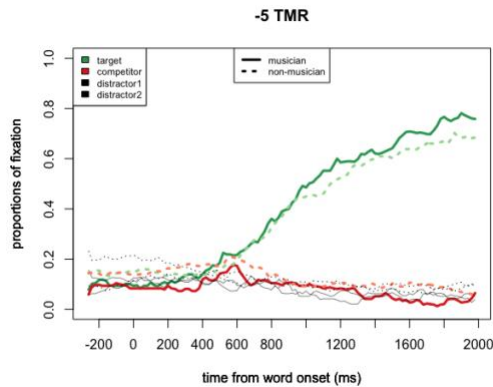


Figure 4. Gaze fixations to the target, competitor and distractor images as a function of time from word onset of the target word in -5 TMR condition.

The event related pupil dilation responses (ERPD) are calculated with the following formula:

$$\% \text{ ERPD} = \frac{\text{observation} - \text{baseline}}{\text{baseline}} * 100$$

The baseline is taken to be the mean of the pupil response occurring 200 ms right before the onset of the target word. The ERPD in response to the processing of the target word reflects that there is less change in pupil size for non-musicians in quiet. Overall, both groups exhibit less change in ERPD in the -5 TMR condition compared to the 0 TMR condition. For the musician group, quiet condition appears to have the biggest change in ERPD across conditions (Figure 5).

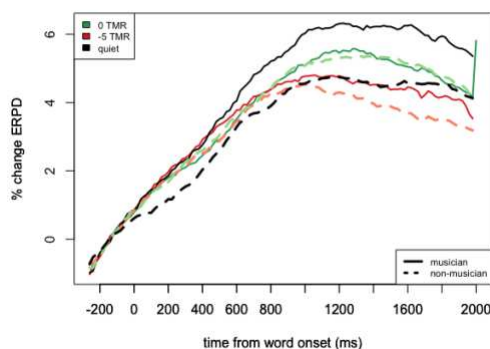


Figure 5. Pupil dilation data time curves for both musicians and non-musicians at all conditions (quiet, 0 TMR, -5 TMR).

Conclusion

The preliminary results so far indicate that musicians may perform slightly better at the sentence recall task as the background speech masker level increases. If this holds as data collection is completed, this finding would be in line with

previous findings, despite the difference in masker structure (two talker vs one talker masker) (Başkent & Gaudrain, 2016).

In the visual world paradigm, lexical competition, i.e. listeners' simultaneous consideration of phonologically similar words to be the potential target, which captures automatic process observed in previous research (Salverda, Dahan, & McQueen, 2003; Wagner et al., 2016) is present in the quiet condition, while it becomes less visible as the masking increases, especially for musicians. The overall timing of lexical decision making is delayed for both groups when the target is presented with the masker. The event related pupil dilations, which may be reflecting either increased effort or more attention allocation are inconclusive. We aim to collect more musician data before drawing any further conclusions and conducting statistical analysis.

Acknowledgements. This project is supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie- Skłodowska-Curie grant agreement No 675324, and the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), VICI Grant (Grant No. 918-17-603). This study has been approved by the Medical Ethics Committee (METc) of University Medical Center Groningen (UMCG). □□

References

- Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, 139(3), EL51-EL56.
- Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, 137(1), 378-387.
- Calandruccio, L., Buss, E., & Bowdrie, K. (2017). Effectiveness of Two-Talker Maskers That Differ in Talker Congruity and Perceptual Similarity to the Target Speech. *Trends in Hearing*, 21, 1-14.
- Coffey, E. B. J., Mogilever, N., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: a review. *Hearing Research*, 352, 49-69.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6(1), 84-107.
- Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, 8(179).
- Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, 1252(1), 100-107.
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician

- enhancement for speech-in-noise. *Ear and Hearing*, 30(6), 653–661.
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133(4), 2431–2443.
- Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of Musical Training on Understanding Voiced and Whispered Speech in Noise, 9(1).
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90(1), 51–89.
- Salverda, A. P., & Tanenhaus, M. K. (2017). The Visual World Paradigm. In *Research methods in psycholinguistics: A practical guide* (pp. 89–110).
- Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. *Cognitive Processing*, 17(1), 79–87.
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, 5(11628), 1–10.
- Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America*, 107(3), 1671–84.
- Wagner, A. E., Toffanin, P., & Baskent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Frontiers in Psychology*, 7(398).